

Spatial Information in Large-Scale Neural Recordings

Thaddeus R. Cybulski^{1,*}, Joshua I. Glaser^{1,*},
Adam H. Marblestone^{2,3}, Bradley M. Zamft⁴,
Edward S. Boyden^{5,6,7}, George M. Church^{2,3,4}, and
Konrad P. Kording^{1,8,9}

¹*Department of Physical Medicine and Rehabilitation, Northwestern University
and Rehabilitation Institute of Chicago, Chicago, Illinois, USA*

²*Biophysics Program, Harvard University, Boston, Massachusetts, USA*

³*Wyss Institute, Harvard University, Boston, Massachusetts, USA*

⁴*Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA*

⁵*Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts,
USA*

⁶*Department of Biological Engineering, Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA*

⁷*McGovern Institute, Massachusetts Institute of Technology, Cambridge,
Massachusetts, USA*

⁸*Department of Physiology, Northwestern University, Chicago, Illinois, USA*

⁹*Department of Applied Mathematics, Northwestern University, Chicago,
Illinois, USA*

* Authors contributed equally to this work. Direct correspondence to
cyb@northwestern.edu and j-glaser@u.northwestern.edu

Abstract

To record from a given neuron, a recording technology must be able to separate the activity of that neuron from the activity of its neighbors. Here, we develop a Fisher information based framework to determine the conditions under which this is feasible for a given technology. This framework combines measurable point spread functions with measurable noise distributions to produce theoretical bounds on the precision with which a recording technology can localize neural activities. If there is sufficient information to uniquely localize neural activities, then a technology will, from an information theoretic perspective, be able to record from these neurons. We (1) describe this framework, and (2) demonstrate its application in model experiments. This method generalizes to many recording devices that

resolve objects in space and should be useful in the design of next-generation scalable neural recording systems.

Contents

1	Introduction	3
2	Framework	4
2.1	Localization and Resolution	4
2.2	Fisher Information: General Principles	6
2.3	Cramer-Rao Bounds	7
2.4	Independence and Summation	7
2.5	Fisher Information: Single-Source Resolution	8
2.6	Fisher Information: Differential Resolution	9
2.7	Point Spread Functions and Signal Intensity Distributions	9
3	Framework Discussion	11
4	Demonstrations	12
4.1	Assumptions	13
4.2	Single Neuron Localization	13
4.3	Electrical Sensing	14
4.4	Optical Sensing	16
4.4.1	General Information	16
4.4.2	Wide-field Fluorescence Microscopy	16
4.4.3	Two-photon Microscopy	19
4.5	Technological Optimization	21
4.6	Technology Comparison	24
5	Demonstrations Discussion	24
6	Additional Methods	27
6.1	Noise Calculations	27
6.2	Applications: Electrode Grid Analysis	27
7	Acknowledgements	28
8	References	29
9	Supplementary Information	35
9.1	Noise Sources	35
9.2	Fisher Information Derivation	37

1 Introduction

A concerted effort is underway to develop technologies for recording simultaneously from a large fraction of neurons in a brain [1,2]. For a technology to reach the goal of large-scale recording, it must gather sufficient information from each neuron to determine its activity. This suggests that neural recording methodologies should be evaluated and compared on information theoretic grounds. Still, no widely applicable framework has been presented that would quantify the amount of information large-scale neural recording architectures are able to capture. Such a framework promises to be useful when we want to compare the prospects of new recording technologies.

A neural recording technology can be judged by its ability to isolate signals from individual neurons. One common method of differentiating between signals from different neurons is through the neurons' locations: if the recording technique can determine the signal sources are sufficiently far apart (by signal amplitude or other methods), then the signals likely came from different neurons. One can quantify this ability to spatially differentiate neurons using Fisher information, which measures how much information a random variable (e.g. a signal on a detector) contains about a parameter of interest (e.g. where the signal originated). Fisher information can be used to determine the optimal precision with which the parameter of interest (the neural location) can be estimated.¹ By calculating the Fisher information a technology carries about sources it records, one can determine how precisely neural locations can be estimated using this technology, and thus whether the neural activities can be distinguished in space.

Determining the Fisher information content of a sensing system allows determining the informatic limits of a technology in a given situation. These informatic limits, in turn, can guide technology design. For example, by quantifying the information content of an electrode array as a function of the spacing between electrodes, one could determine the spacing necessary to distinguish neural activities. Similarly, one can compare the information content of several optical recording approaches to determine the optimal technology for a given experiment.

Here we develop a Fisher information-based framework that characterizes neural recording technologies based on their abilities to distinguish activities from multiple neurons. We apply this framework to models of neural recording techniques, describe how the Fisher information scales with

¹Fisher information is a theoretical calculation that determines the best a technology can do – signal separation techniques (e.g. [3]) are generally required to approach this optimum.

respect to recording geometries and other parameters, and demonstrate how this framework could be utilized to optimize experimental design. We demonstrate the utility of a Fisher information-based evaluation of neural recording technologies, which may inform the design and development of next-generation recording techniques.

2 Framework

2.1 Localization and Resolution

A fundamental concern in neural recording is localization, the ability to accurately estimate the location of origin of neural activity. Localization is a primary method of determining the identity of an active neuron.

The problem of establishing neural locations can be split into two separate regimes. One regime is when an active neuron has no active neighbors (Figure 1A). In this state, we are chiefly concerned with the ability to attribute the signal to the correct neuron (single-source resolution [4]). This can be done by accurately localizing one activity at a given time on a background of noise (Figure 1B). The other regime is when two neighboring neurons are simultaneously active (Figure 1C). In this state, we are chiefly concerned with the ability to differentiate the two neurons, i.e. are there two clearly distinguished or one blurred neuron (differential resolution [4]). This can be done by simultaneously localizing the activities of both neurons accurately (Figure 1D).²

Fisher information can be used to determine whether both scenarios are theoretically possible for a given technology. Here we treat both of these scenarios: first by calculating the Fisher information a sensing apparatus has about the location of a single neuron, then expanding this framework to treat location parameters of multiple neurons. We address localization and resolution in the theoretical limit where the point spread function (PSF) is known, in order to study the limiting effects of neuronal and sensor noise

²While we have been discussing differentiating neurons, the framework itself differentiates between point sources. In this paper, we make the assumption that separate point sources belong to separate neurons. In reality, it is possible that there could be separate signals from the cell body and dendrites that are perceived as different sources. These can be united using additional information (e.g. anatomical imaging or simultaneous activity).

on localization precision.³

Regardless of the number of neurons and sensors we are treating, Fisher information gives us a metric with which to evaluate a recording technology. Spatial information, the amount of information regarding the location of a source (i.e., a quantitative measure of localization ability), can be used to determine whether it is possible to correctly attribute an activity to its source (or multiple activities to multiple sources). In order to know the identity of a source, we must be confident about the location of origin of the activity with a positional error less than δ , where δ is the distance from one neuron to another (Figure 1B&D). In terms of Fisher information, if we have sufficient information to locate the source of activity with a precision δ , we can assign that activity to a single neuron that occupies that location.

³There exists a family of deconvolution techniques that estimate the PSF and use it to obtain a more accurate representation of the original signal (e.g. [5–8]). In theory, with sufficient samples and knowledge of the PSF, one could obtain a perfect representation of a sparse signal in the absence of noise. This is not the case in practice, as signals are not only modified reversibly by PSFs, but are modified irreversibly by noise on neurons and detectors (e.g. [9, 10]). In the presence of noise and other aberrations, it thus becomes difficult to isolate individual sources using deconvolution techniques, even when the PSF is known. Thus, it is interesting to determine the isolated effects of noise on recording methods. Moreover, as this Fisher information framework gives optimal bounds on precision with a known PSF, it can be used to determine how close to optimal a deconvolution algorithm performs.

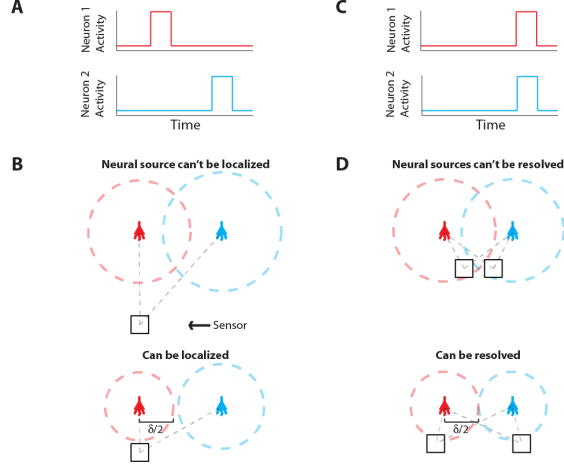


Figure 1: Localization and Resolution. (A) In many behavioral states, neural systems have sparse activity, in which neighboring neurons (red and blue) are not active at the same time. In this scenario of single-source resolution, one neuron must be localized at a given time. Panel B looks at this scenario. (B) Two neighboring neurons are shown a distance δ away from each other. Dotted lines indicate regions where we are confident about the source of a signal, i.e. we have a sufficient amount of information regarding that signal's location. The signals from the two neurons are recorded by the sensor at different times and do not interfere with each other. When a neuron cannot be localized effectively, i.e. there is not sufficient Fisher information, it is because the signal from that neuron was not strong enough to overcome noise. (C) Sometimes, neighboring neurons are simultaneously active. In this scenario of differential resolution, both neurons must be localized at a given time. Panel D looks at this scenario. (D) Same as B, except two sensors are necessary for differential resolution. When both sensors record similar signals, i.e. when there is large mutual information regarding the two neurons' activities, it is difficult to resolve the neurons.

2.2 Fisher Information: General Principles

Fisher information is a metric that measures the information a random variable has about a parameter, and can be used to determine how well that parameter can be estimated. More precisely, Fisher information, $\mathcal{I}(\theta)$ is a measure of the information a random variable X , with distribution

$f(X; \theta)$ parameterized by θ , contains about the parameter θ [11]:

$$\mathcal{I}(\theta) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \log f(X; \theta) \right)^2 \middle| \theta \right] = \int \left(\frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 f(x; \theta) dx \quad (1)$$

Intuitively, the more X changes for a given change in θ (the greater the magnitude of $\frac{\partial}{\partial \theta} f(X; \theta)$), the more information you will know about θ by observing X .

More generally, the Fisher Information a random variable X has about a parameter vector $\boldsymbol{\theta}$ with k elements $[\theta_1 \cdots \theta_k]$ can be represented by a $k \times k$ matrix with elements:

$$(\mathcal{I}(\boldsymbol{\theta}))_{ij} = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_i} \log f(X; \boldsymbol{\theta}) \right) \left(\frac{\partial}{\partial \theta_j} \log f(X; \boldsymbol{\theta}) \right) \middle| \boldsymbol{\theta} \right] \quad (2)$$

The elements of this matrix represent the information contained in a sample about a pair of parameters.

2.3 Cramer-Rao Bounds

The optimal precision with which the parameter, θ , can be estimated is inversely related to the Fisher information contained about that parameter. More precisely, the variance of an unbiased estimator of a parameter is lower bounded by the Cramer-Rao bound (CRB) [12]:

$$\text{Var} [\hat{\boldsymbol{\theta}}_i] \geq [\mathcal{I}(\boldsymbol{\theta})^{-1}]_{ii} \quad (3)$$

An important implication of this is that the CRB on θ_i not only depends on the information X contains about θ_i , but how similar θ_i 's effect on X is to the rest of the elements of $\boldsymbol{\theta}$. An off-diagonal term $(\mathcal{I}(\boldsymbol{\theta}))_{ij}$ with large magnitude means that the parameters i and j are strongly correlated (or anti-correlated) in terms of their input on X . This will increase the CRB on estimating parameters i and j .

2.4 Independence and Summation

If two observations X_1 and X_2 are independently affected by $\boldsymbol{\theta}$, then the two Fisher information matrices about $\boldsymbol{\theta}$ can be summed, as could be expected

by the implications of independence on sample variance. This property allows us to easily apply our framework to situations with multiple samples, either by multiple sensors or multiple time points.

In the following sections, we will apply the above properties of Fisher information and CRBs to develop a framework for determining how precisely the location of neural activities can be estimated, and thus whether they can be distinguished. Note that, while we will describe the ability to distinguish neurons solely using spatial information, additional sources of information can be used, e.g., temporal information in optical [13] and electrical recordings [14] (see *Demonstrations Discussion*).

2.5 Fisher Information: Single-Source Resolution

We first examine the situation where a single active source of some known intensity must be localized using an ensemble of sensors.⁴ Here we observe a random variable, X , the value recorded at some sensor (e.g. in Volts). $f(X; \theta)$ then is the distribution of sensor values from repeated recordings of a neuron parametrized by θ . θ is a vector representing spatial (and other) parameters that characterize the neural signal. This resulting distribution $f(X; \theta)$ reflects both intrinsic variance of a neural signal as well as extrinsic factors such as other neurons and noise.

Here, Fisher information, $\mathcal{I}(\theta)$, measures how much the distribution of recorded sensor values $f(X; \theta)$ tells us about the location of a signal's origin (Figure 2D). Intuitively, if a change in the signal origin's location would cause a large change in the recorded signal, then there will be a large amount of information about the location. However, if a change in the origin of the neural signal does not affect the recorded signal, there will be little information about the location of the neuron.

The CRB for a given parameter θ_i will tell us how precisely that location parameter can be estimated from the signal intensity. Assuming an unbiased estimator (the average estimate will be the true location), the best possible variance of the estimate is $[\mathcal{I}(\theta)^{-1}]_{ii}$. If we want to be confident that the estimated location of a given neuron's activity is within $\delta/2$ of its true location, as in Figure 1, the CRB on the estimate of distance must be less than $(\delta/4)^2$.⁵

⁴Activity in neural systems is often sparse [15–19]; this simplified scenario may be a useful model of neural systems.

⁵95% confidence under Gaussian assumptions

Without assuming any prior knowledge, at least k variables are required to estimate k parameters, as the system is underconstrained with smaller numbers of samples. In our case, we need multiple sensors in order to estimate a neuron’s location. If the sensors have independent noise – an assumption we use in our demonstrations – the information matrices can be summed (See *Independence and Summation*).

2.6 Fisher Information: Differential Resolution

In the scenario of multiple neurons acting simultaneously, we are interested in using signals recorded from an ensemble of sensors to estimate the location parameters of each neuron. That is, $\boldsymbol{\theta}$ now represents the location parameters of all neurons in the system, and $f(X; \boldsymbol{\theta})$ represents the distribution of signal intensities on a sensor given all of the neurons in the system. We can then construct a Fisher information matrix to determine the precision with which each parameter can be estimated. If each sensor recording is affected by n neurons, each with k parameters, the Fisher information matrix will be $nk \times nk$. The CRB calculated in this scenario will be most applicable to determining whether technologies are able to effectively record from a population of neurons.

2.7 Point Spread Functions and Signal Intensity Distributions

To determine the spatial Fisher information, we must know the distribution of signals on a sensor given the location of the activity, $f(X; \boldsymbol{\theta})$. In this section, we derive the general form of $f(X; \boldsymbol{\theta})$ based on the PSF of a technology.

The signal measured by many recording systems is well approximated as a linear function of the signals from each neuron in a population [20, 21], i.e. the total sensor signal is the sum of the individual neural signals weighted by the magnitude of their individual effects on the sensor (Figure 2A&B). We thus only consider linear interactions; it should be noted that the Fisher information framework is also compatible with nonlinear interactions (e.g. sensor saturation). For N neurons and M sensors in a system, in the absence of noise, the signal on any particular sensor can therefore be described as:

$$\mathbf{x} = \mathbf{W}\mathbf{a} + \epsilon \tag{4a}$$

where \mathbf{x} is the vector of signals on sensors $[X_1, \dots, X_M]$, ϵ is noise on the signal from neurons and sensors, and \mathbf{a} is the vector of signals from neural activities, $[I_1, \dots, I_N]^T$, e.g. the fluorescent signal produced due to neural activity in optical techniques or the voltage signal in electrical techniques. \mathbf{W} is the matrix of PSFs:

$$\mathbf{W} = \begin{bmatrix} w(\mathbf{d}^{1,1}) & \dots & w(\mathbf{d}^{1,N}) \\ \vdots & \ddots & \vdots \\ w(\mathbf{d}^{M,1}) & \dots & w(\mathbf{d}^{M,N}) \end{bmatrix} \quad (4b)$$

where w is the PSF, which depends on the location of the neuron relative to the sensor and other parameters of a recording modality (e.g. light scattering). $\mathbf{d}^{i,j}$ is a vector that gives the location of neuron j relative to sensor i . It has elements $[d_1^{i,j} \dots]$ that describe single location parameters of $\mathbf{d}^{i,j}$.

Combing Eqs. 4a and 4b, we can write the total signal on a sensor i as

$$X_i = \sum_j I_j w(\mathbf{d}^{i,j}) + \epsilon \quad (5a)$$

We can write a function $f(X_i)$ that characterizes the distribution of signal intensities on a sensor. Here, we assume that the noise, ϵ , can be approximated by a zero-mean Gaussian with variance σ_{noise}^2 , so that:

$$f(X_i; \boldsymbol{\theta}) = \mathcal{N} \left(\sum_j I_j w(\mathbf{d}^{i,j}), \sigma_{noise}^2 \right) \quad (5b)$$

where $\mathcal{N}(\mu, \sigma^2)$ signifies a normal distribution (Figure 2C). $\boldsymbol{\theta}$ is the vector of parameters that we are estimating. It can include any I_j and any elements of any $\mathbf{d}^{i,j}$. This allows us to calculate the Fisher information in signal X_i about location parameters of neurons using Eqs. 1 or 2 (Figure 2D).

It is important to note that, as long as they can be analytically described, all types of noise (of which there are many; see *Supplementary Information* for further discussion) can be incorporated into this framework. This flexibility in noise sources makes this framework especially relevant for neural recording.

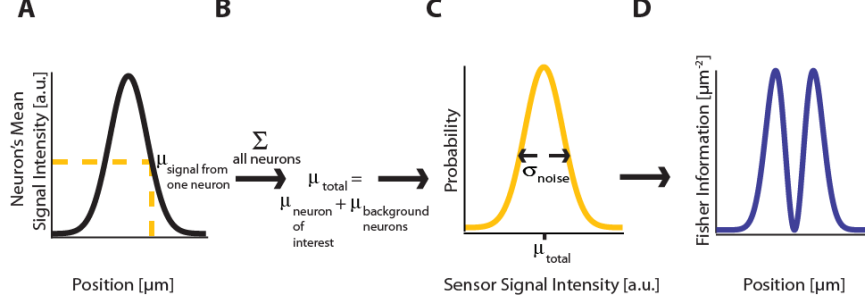


Figure 2: Fisher Information. (A) A signal on sensor i from a neuron j at a particular location has a mean intensity, defined by a recording method’s point spread function and the intensity of the signal from the active neuron. We here plot this mean signal intensity as a function of one position parameter. (B) The mean total signal on a sensor, μ_{total} , is the sum of the signals from every neuron. (C) The distribution of intensities recorded on a sensor is a function of the total mean signal, μ_{total} , and the variance of that signal, σ_{noise}^2 , which can result from many different noise sources. (D) Fisher information can be derived from the distribution of signal intensity values on a sensor.

3 Framework Discussion

Here we have described a framework to quantitatively approach the challenges of large-scale neural recording and determine the necessary experimental parameters for potential recording modalities. This framework extends previous work applying Fisher information to individual imaging techniques (e.g. [10, 22–31]) by considering a PSF and noise model based on recording in neural tissue, and then using the CRB to establish signal separability. It is able to describe the information content of neural recording technologies that separate sources based on their locations. This information content can then be used to evaluate a technology’s ability to separate sources. Such a framework promises to be useful in evaluating and comparing novel and established recording technologies.

Given this framework’s reliance on signal modulation by PSFs, it neglects other ways that sources can be separated, such as color [32] or spike waveform. Some of this information could be made compatible with our framework via virtual recording channels, e.g. in time. While these types of non-spatial information are not considered here, they may be necessary to

separate sources under certain recording situations, e.g. where the dendrites of one neuron produce a signal within the CRB of the cell body of another neuron. In an extreme case, proposed intracellular molecular recording devices have no spatial information, but could still effectively separate signals. [33, 34] While spatial Fisher information is an attractive method of evaluating neural recording techniques, it is important to remember these limitations when considering non-spatial techniques.

In addition, the CRBs described here only consider unbiased estimators. That is, they only provide a lower bound on localization ability when there are no prior assumptions about neurons locations. It is possible to be more precise than the CRB if the estimator is biased (i.e. if assumptions are made about neurons locations, or neurons locations are constrained). There is work on Bayesian Cramer Rao Bounds [35, 36] and bounds on parameter estimation with constraints [37, 38] that could be applied to better understand the capabilities of recording technologies.

This framework is particularly suited to the evaluation of novel techniques due to its general nature; it is applicable to any technique where a spatial PSF can be measured and the system’s noise distribution can be either modeled or explicitly described. For instance, advanced optical techniques, [39, 40] ultrasound, and MRI have all been proposed as potential large-scale neural recording techniques [2, 41]. With a PSF describing how signals from different positions in the brain reach a sensor (some discussion in [40, 42–46]) and further quantification of recording noise, this framework could easily be applied to determine bounds on signal separability for those techniques.

Ultimately, the utility of this approach is dependent on the quality of PSFs and noise models we have. For some techniques, these are well-described (especially PSFs); for others, these are poorly understood. As models of neural recording techniques advance, the predictions of this technique will become more accurate.

4 Demonstrations

Here, we demonstrate the utility of the Fisher information framework for analysis of neural recording technologies. We provide demonstrations of the use of Fisher information in the cases of single-source and differential resolution. We first calculate the spatial Fisher information of a single source in simple recording setups for several model recording methods. We next

demonstrate more realistic uses of the Fisher information framework using multiple neurons: optimal technology design and technology comparison.

4.1 Assumptions

For our demonstrations, we make several assumptions. First, we assume that all activity from the neuron of interest, including the noise, is part of the signal of interest. Thus, the total noise is a function of the sensor noise plus the noise of all neurons except for the neuron of interest. In order to create an accurate model of a neural recording technology, we must know how all sources of noise affect the recorded signal, and also the relation between the noise and the intensity of the neural activity. Because these are in general not known, we make further assumptions in our simulations.

In regards to neural activities, we assume that every active neuron has the same activity I_0 , while non-active neurons have no activity, that the neuron of interest, k , is active at the moment we sample, and that other neurons are active at a uniform rate. We assume noise sources from neurons are independent, so that:

$$\sigma_{noise}^2 = \sum_{j \neq k} \sigma_j^2 \quad (6)$$

There are many sources of noise, both on neurons and sensors, that could be included; these are discussed in the *Supplementary Information*. For our demonstrations, we consider signal dependent noise that can arise from neurons and/or sensors. Specifically, for analytic simplicity, we only consider noise that has a standard deviation proportional to the mean signal: $\sigma_j^2 \propto I_0^2 (w(\mathbf{d}^{i,j}))^2$. We use these simplifying assumptions so that the magnitudes of the fluorescence (optical) and waveform voltage (electrical) have no influence on the final information theory calculations (and the relationship between these magnitudes and the noise is not in general well understood). We emphasize that these simulation assumptions are implemented to simply demonstrate the use of this framework; more realistic outputs could be found using more complex, realistic noise models.

4.2 Single Neuron Localization

Here we calculate Fisher information of recording technologies using a single neuron and simple sensor arrangements as an illustration of our framework.

We look at three technologies: (1) electrical recording, a traditional neural recording modality, (2) wide-field fluorescence microscopy, a traditional optical approach, and (3) two-photon microscopy, a modern optical approach. These examples are chosen for their relative simplicity and ability to illustrate the flexibility of a Fisher information approach to modeling neural recording.

For any technology, the aim is for there to be, across all sensors, sufficient information about every location in the brain in order to identify a neuron firing in that location. Thus for an individual sensor, it can be better to have sufficient (enough to identify a neuron, as in Figure 1) information spread over a large area than excessive information about a small area. This suggests that experimental designs could be modified to get sufficient information for the required task. For example, an optical technology may have extra information at low depths, but insufficient information at large depths. In this case, the PSF could be modulated (e.g. [22]) to decrease low-depth information (making those images blurrier), while increasing high-depth information.

4.3 Electrical Sensing

The electrical potential from an isolated firing neuron decays approximately exponentially with increasing distance [47, 48], at least at short distances. Here, we model a simple electrical system: an isotropic electrode with spherical symmetry. In this isotropic approximation, the PSF has an exponential decay with radial distance from the electrode tip (PSF taken from Table 1, using parameters found in Table 2; Figure 3B).

For electrical recording, estimators of location parameters have the lowest standard deviation σ_x and σ_y when in-between two electrodes, and the lowest σ_z when directly above or below an electrode (Figure 3D&E). Generally, we see that electrical recordings provide relatively weak information over a relatively wide area. In fact, we find that, in “worst-case” regions, standard electrode arrays should have difficulty localizing a source within the bounds required to discriminate between neighboring neurons. Given that current arrays generally require more information than a single sample of signal intensity to sort spikes (e.g. waveform shape is used), this is an expected result.

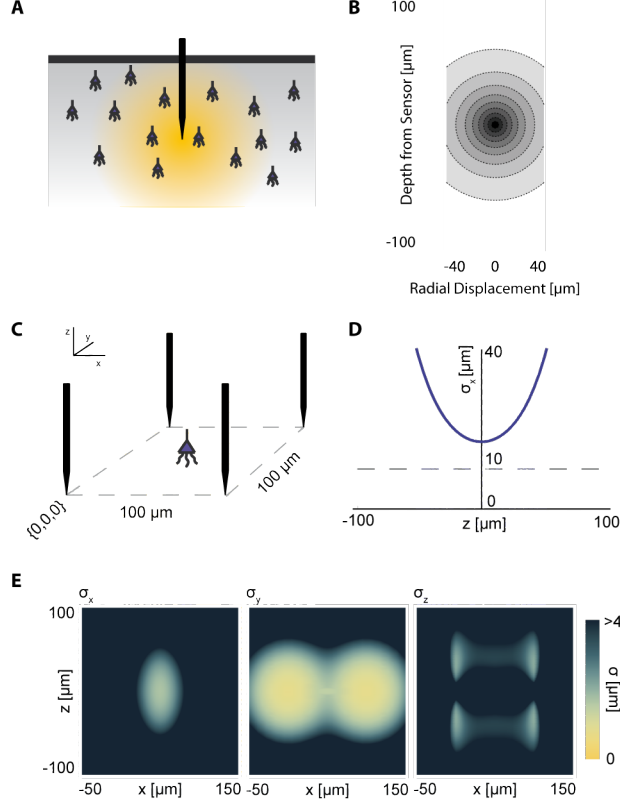


Figure 3: Electrical Recording. An overview of the modeling and Fisher information analysis of electrical recording. **(A)** Schematic: An electrode records electrical signals directly from nearby neurons. **(B)** The spatial PSF for a single electrode recording, valued in arbitrary units, for an electrode located at (0,0,0). **(C)** A schematic for the simple 4-electrode recording system simulated here. Electrodes are arranged in a $100\mu\text{m} \times 100\mu\text{m}$ square, all with $z = 0$. The coordinate system for **(D)** and **(E)** is defined. **(D)** The standard deviation of an estimator for position on the x axis (σ_x) for a source located at $(50, 50, z)$. The grey dashed line indicates a CRB standard deviation of $10\mu\text{m}$. This $10\mu\text{m}$ standard deviation corresponds to a 95% accuracy of determining the correct active neuron for neurons whose centers are $40\mu\text{m}$ apart, and assuming a Gaussian estimation profile. **(E)** Standard deviation of estimators for x , y , and z location (σ_x , σ_y , σ_z) for a source located at $(x, 50, z)$. See Table 1 and Table 2 for equations and parameters used to generate this figure.

4.4 Optical Sensing

4.4.1 General Information

Optical recording of neural activity generally relies on fluorescent dyes that are sensitive to activity. In order to measure this signal, a neuron must be illuminated with light in the dye’s excitation spectrum. Light is then emitted by the dye at a distinct, longer (lower energy) wavelength, which is picked up by a photodetector. Optical signal transmission is subject to absorption, scattering, and diffraction, which degrade the emitted signals with distance. Absorption of light effectively cause an exponential decrease in intensity of detected photons as light travels through a medium [49, 50]. Scattering can affect light in multiple ways; high-angle scattering diverts photons from the detector and produces an effect similar to absorption, while low-angle scattering causes blurring of the image on the detector. This blurring increases approximately linearly with depth into the tissue [51]. Finally, diffraction results when light passes through an aperture, creating the finite-width Airy disk [52]. In our optical PSFs, we assume scattering and diffraction result in Gaussian blurring [51, 53]. Our PSFs assume imaging through a single homogeneous medium; in practice, tissue inhomogeneity and refractive index mismatch can produce additional aberrations in the absorption, scattering, and diffraction domains that we do not model here.

In a typical optical setup, a lens focuses a set of photons from one point in space onto a corresponding point behind the lens. This phenomenon can be used either to focus incident light onto a desired location for illumination, or to focus emitted light from the focal plane onto a photodetector for imaging. Photons from outside the focal plane will be blurred, and this blurring increases linearly as distance from a focus point increases [54, 55]. We also assume defocusing results in Gaussian blurring [54, 55].

4.4.2 Wide-field Fluorescence Microscopy

Neural activity in a focused optical system is generally sensed using fluorescent dyes, which require some excitatory light. In the canonical optical example of wide-field microscopy, an entire volume is illuminated (Figure 4A). The PSF for this technology takes the above effects of absorption, scattering, diffraction, and defocusing into account; we assume total illumination so that the PSF here models the spread of the emission light (Figure 4B, PSF taken from Table 1 using parameters found in Table 2).

For optical recording with a simple lens, estimators of location parameters

have lowest standard deviation σ_x , σ_y , and σ_z when centered above the imaging system in the focal plane (Figure 4D&E). For large depth, the ability to distinguish locations decreases rapidly due to photon loss caused by scattering and absorption (Figure 4D&E). For medium depth ranges, scattering blurs the image, even on the focal plane. These phenomena decrease the utility of deep focal-plane wide-field optics in tissue. At shallower focal depths, optical recordings provide a large amount of information on the focal plane, while carrying relatively little information about sources out of the focal plane (Figure 4D&E).

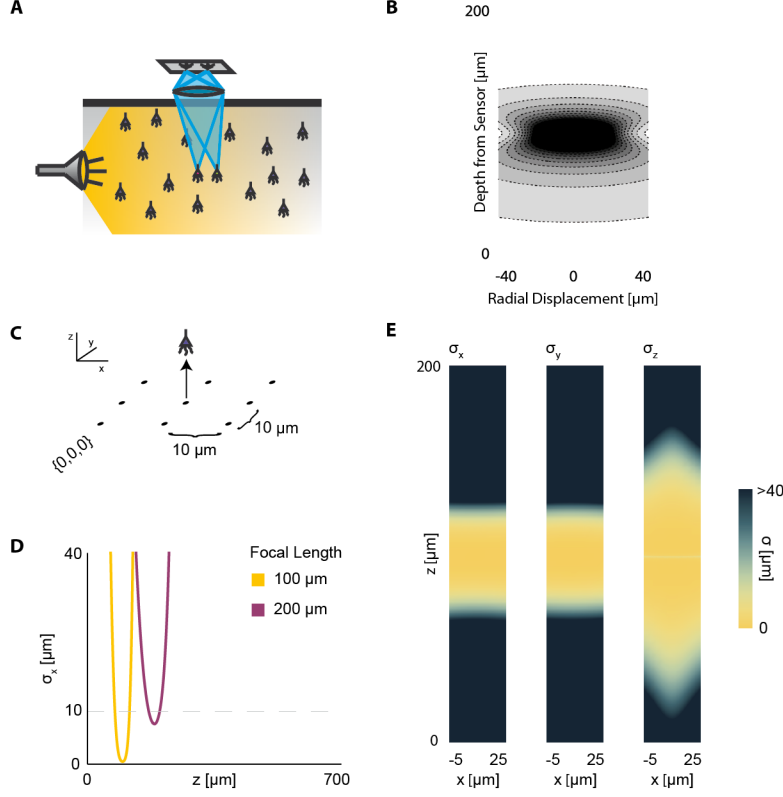


Figure 4: Wide-field Fluorescence Optical Recording. An overview of the modeling and Fisher information analysis of wide-field fluorescence optical recording. **(A)** Schematic: The whole recording volume is illuminated; dye in active neurons fluoresces and emits light; the emitted light is focused by a lens onto a photosensor. **(B)** The spatial PSF for wide-field fluorescence optical recording, valued in arbitrary units, for a lens centered at (0,0,0) with a focal plane at 100 μm. **(C)** A schematic for the simple 9-sensor optical recording system simulated here. Sensors are arranged in a 3×3 grid with a pitch of 10 μm, all sensors with $z = 0$. The coordinate system for **(D)** and **(E)** is defined. **(D)** The standard deviation of an estimator for position on the x axis (σ_x) for a source located at (10, 10, z) and an optical system with focal depth of either 100 μm or 200 μm. The grey dashed line indicates a CRB standard deviation of 10 μm. **(E)** Standard deviation of estimators for x , y , and z location (σ_x , σ_y , σ_z) for a source located at (x , 10, z) and an optical system with focal depth of 100 μm. See Table 1 and Table 2 for equations and parameters used to generate this figure.

4.4.3 Two-photon Microscopy

In two-photon microscopy, long-wavelength incident light (i.e. composed of low-energy photons) is focused onto a single point of interest to excite fluorophores in that area. In order for the fluorophore to emit light, two low-energy photons must be absorbed nearly simultaneously; the likelihood of this event is proportional to the square of the intensity of incident light at a point. Effectively, this concentrates the area of sufficient illumination to a volume nearby the focal point of the incident beam (while increasing the illumination power requirements) [56]. Like with wide-field fluorescence microscopy, the PSF is a function of defocusing, absorption, and scattering (Figure 5B, PSF taken from Table 1 using parameters found in Table 2). We assume total photon capture so that the PSF here models the spread of the excitation light.

For two-photon microscopy, estimators of location parameters have lowest standard deviation σ_x , σ_y , and σ_z just above and below the focal plane (Figure 5C). Perhaps counter-intuitively, there are extremely-high or undefined σ 's along the focal plane. This is due to our simplified recording setup: given the tightly-focused PSF for two-photon microscopy, sources very close to the focal plane of our setup are effectively only “seen” by one sensor. Thus, we cannot gather meaningful information about the source’s three location parameters, resulting in a singular or near-singular Fisher information matrix. In practice, this is alleviated by either decreasing the pitch of sensed regions or applying magnification to the sample, which we do not model here. We also see a reduced dependence on focal depth when compared to a wide-field imaging setting, as expected.

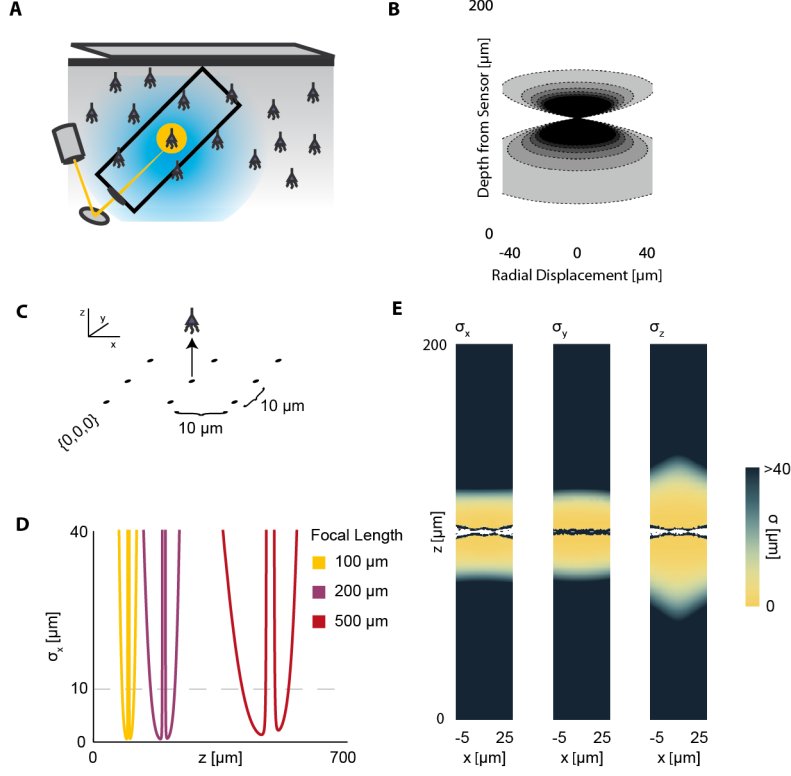


Figure 5: Two-photon Optical Recording. An overview of the modeling and Fisher information analysis of 2-photon optical recording. **(A)** Schematic: incident light is focused onto a particular location in a volume; dye in neurons illuminated by the incident light fluoresces and emits light; the emitted light is sensed by a large single photosensor. The black box indicates the space represented in **B**, with zero depth being located at the lens and increasing depth indicating increasing distance into the brain. **(B)** The spatial PSF for incident light relative to its source in 2-photon optical recording. It is valued in arbitrary units for a lens centered at $(0,0)$ with a focal plane at $100 \mu\text{m}$. **(C)** A schematic for the simple 9-pixel two-photon recording system simulated here. Sampled points are arranged in a 3×3 grid with a pitch of $10 \mu\text{m}$, all points with $z = 0$. The coordinate system for **(D)** and **(E)** is defined. **(D)** The standard deviation of an estimator for position on the x axis (σ_x) for a source located at $(10, 10, z)$ and an optical system with focal depth of $100 \mu\text{m}$, $200 \mu\text{m}$, or $500 \mu\text{m}$. The grey dashed line indicates a CRB standard deviation of $10 \mu\text{m}$. **(E)** Standard deviation of estimators for x , y , and z location (σ_x , σ_y , σ_z) for a source located at $(x, 10, z)$ and an optical system with focal depth of $100 \mu\text{m}$. White regions indicate regions where the Fisher information matrix is ill-conditioned. See Table 1 and Table 2 for equations and parameters used to generate this figure.

Parameter	Value
C_{el}	28 μm [47, 48]
D_{lens}	300 μm (within current dimensions)
λ (wide-field)	633 nm (visible light)
λ (2-photon)	800 nm
γ (wide-field)	0.15 [51, 57]
C_{op} (wide-field)	100 μm (with 515 nm light) [50]
γ (2-photon)	0.002 (with 725 nm light) [58]
C_{op} (2-photon)	200 μm (with 909 nm light) [50]

Table 2: Simulation Parameter Values

Electrical	$w_{el}(r) = \exp\left(\frac{-r}{C_{el}}\right)$
Optical: Wide-field fluorescence microscopy	$w_{wf}(\ell, z) = \frac{Q}{2\pi} \exp\left(\frac{-z}{C_{op}}\right) \frac{1}{(s_{defocus}^2 + s_{dif}^2 + s_{scat}^2)}$ $\times \exp\left(\frac{-\ell^2}{2(s_{defocus}^2 + s_{dif}^2 + s_{scat}^2)}\right)$ $s_{defocus} = \frac{D_{lens}(z_0 - z)}{2z_0}, s_{dif} = \frac{0.42\lambda z}{D_{lens}}, s_{scat} = \gamma z$
Optical: 2-photon microscopy	$w_{2P}(\ell, z) = \frac{1}{\pi} \frac{1}{(s_{defocus}^2 + s_{dif}^2 + s_{scat}^2)}$ $\times \left(Q \exp\left(\frac{-z}{C_{op}}\right) \exp\left(\frac{-\ell^2}{2(s_{defocus}^2 + s_{dif}^2 + s_{scat}^2)}\right) \right)^2$

Table 1: Point Spread Functions of Recording Modalities. Analytic expressions are given for PSFs. r is the distance in any radial direction from the electrode, and ℓ is the lateral distance from the center of the lens for optical techniques. Note that $r^2 = x^2 + y^2 + z^2$ and $\ell^2 = x^2 + y^2$. C_{el} is the spatial constant of electrical decay. C_{op} is the spatial constant of optical decay. $s_{defocus}^2$, s_{dif}^2 , and s_{scat}^2 are the variance of the spread of optical light due to defocusing, diffraction, and scattering, respectively. D_{lens} is the diameter of a lens. λ is the wavelength of the light. z_0 is the focus depth, and Q is the light flux (area per photon).

4.5 Technological Optimization

This example will demonstrate the ability to use Fisher information to ask questions about the necessary experimental parameters of neural recording

technologies. In particular, we will use Fisher information to examine sensor placement in electrical recording. In order to successfully record activity from every neuron in a volume, we must place sensors so that they extract sufficient information about every neural location in that volume. That is, the CRB regarding the ability to estimate the location of each point in a volume must be below some threshold for localization.

Here, we simulate several possible arrangements of electrical sensors and evaluate the information that these systems provide about different locations in a volume. Specifically, we look at five electrode arrangements: (1) electrodes evenly distributed in an equilateral grid (Grid electrodes); (2) randomly placed electrodes (Random electrodes); (3) electrodes evenly distributed in a plane (Planar electrodes); and (4&5) two arrangements of columns of electrodes, where electrodes are densely packed within a column, and these columns are arranged in a grid [59] (Column electrodes) (Figure 6A). Here, we assume that noise is independent between sensors, i.e. noise is all on the sensor. Under this assumption, each electrode takes an independent sample of a signal; information about the location of the source of that signal is then additive across sensors. Fisher information here is thus the information the entire ensemble of electrodes provides about a point. In this simplified example, we determine localization, rather than resolution, capabilities, which corresponds to the common situation of sparse neural firing. Multiple sources would necessarily reduce the amount of information contained about individual sources and would be geometry-dependent.

In this simplified simulation, Grid electrodes and Random electrodes have the best performance, as they sample space uniformly (Grid) or almost uniformly (Random) (Figure 6B). Due to the regular nature of Grid electrodes, there is the added benefit of a guaranteed lower bound for information carried about locations in a volume. Planar electrodes are able to estimate a small fraction of locations very well, but carry very little information about most locations in a volume. Columnar electrodes, in general, have the interesting property that the z coordinate can be estimated more accurately, due to the density of electrodes in this direction. It’s also important to note that the feasibility of Columnar electrodes will likely depend on the spacing between shanks. As the shanks move closer together (e.g. the bottom row compared to the fourth row), a greater number of neurons will be able to be distinguished. The use of this Fisher information framework promises to inform sensor placement decisions.

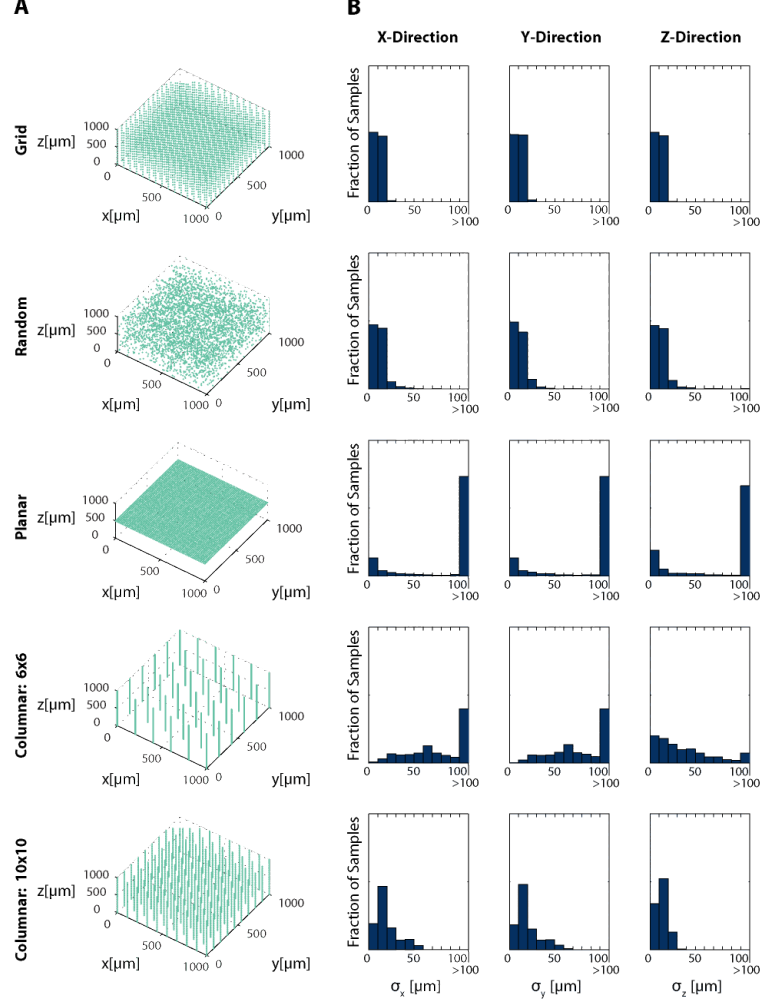


Figure 6: Electrode Placement and Fisher Information. CRBs on the x, y, and z coordinates of neurons using various electrode arrays. We simulate $\sim 3.5 \times 10^3$ electrodes in a $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ cube of brain tissue. Electrodes were arranged in one of five patterns: uniformly distributed in a grid throughout the volume (top row), random placement (second row), electrodes uniformly distributed on a plane at $500 \mu\text{m}$ depth (third row), a 6×6 grid of columns of electrodes with 100 electrodes evenly distributed in each column (fourth row), and a 10×10 grid of columns of electrodes with 30 electrodes evenly distributed in each column (bottom row). Total Fisher information about a point consists of the sum of information contained about that point in each sensor. **(A)** Distribution of electrodes in the volume for each pattern. **(B)** Distribution of Cramer-Rao bounds about a random sample of 10^4 points in the volume. Standard distributions are shown. The three columns represent estimation about the x, y, and z coordinates, from left to right. See Table 2 for parameter values.

4.6 Technology Comparison

Finally, we demonstrate the use of Fisher information for determining resolution ability. This example will demonstrate the ability to use Fisher information to compare technologies. In order to determine appropriate technologies for a given situation, it is necessary to know which technology will maximize the information output, and where information will be concentrated for a given technology.

Here we apply this Fisher information framework to a two-source, multi-sensor setup for both wide-field fluorescence and two-photon microscopy in order to determine performance over depth (Figure 7). We find, perhaps confirming intuition, that wide-field and two-photon fluorescence perform similarly for shallow sections, but performance of wide-field fluorescence microscopy degrades significantly at a depth of 500 μm while two-photon performs well at this depth. Interestingly, both methods contain a large amount of information not only about signals near the focal point, but also about sources nearby the lens. This implies that signals could be recovered from out-of-focus samples given proper recording conditions. While this demonstration yielded the expected results, this framework could be used to compare existing technologies in novel situations, or to compare novel technologies.

5 Demonstrations Discussion

We have demonstrated how the Fisher information framework can be applied to neural recording technologies, and have demonstrated possible applications of this framework including determining optimal technology design and comparing technologies under differing recording conditions. In these demonstrations, interesting findings emerged, some of which confirm experimental knowledge. For instance, (1) when using columnar electrodes, increasing the spacing between electrode shanks leads to a very large fall-off in the number of neurons that can be recorded. (2) For shallow recording depths, wide-field and two-photon microscopy have similar performance capabilities, but at larger depths two-photon microscopy becomes significantly better.

We made several simplifications regarding neural activity, noise, and recording technologies when demonstrating the use of the Fisher information framework. However, these approximations were useful in demonstrating a unifying view over recording methodologies in a single paper. Moreover,

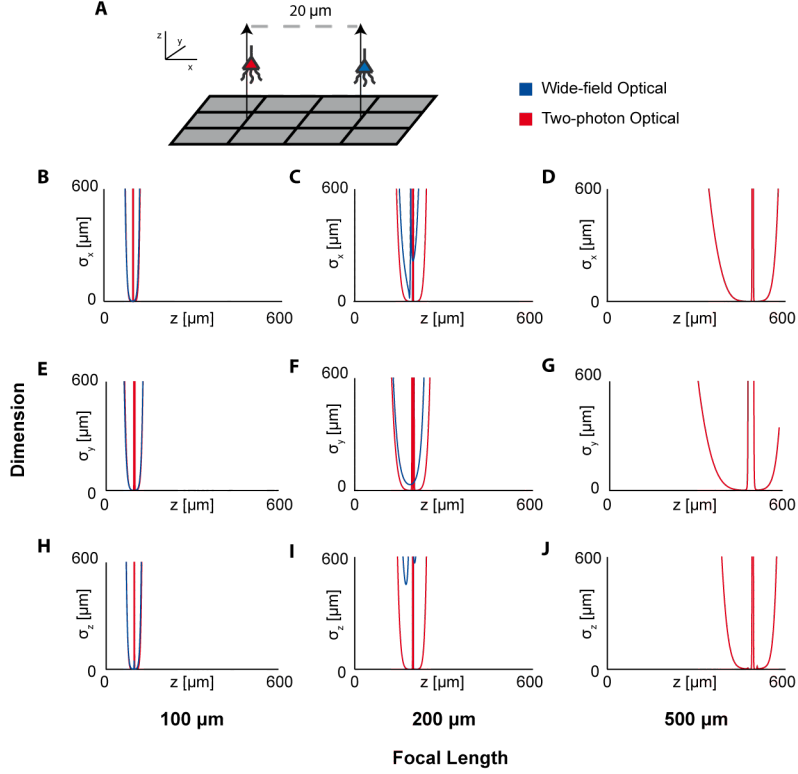


Figure 7: Optical Technology Comparison at Multiple Focal Depths. CRB on the location of the x , y , and z coordinates of a source in a multi-sensor, two-source system. The depth of the sources is varied by an equal amount and the CRB on each of the sources is calculated at each depth (the CRBs of only one source is shown; they are equivalent due to the symmetric setup). This analysis is performed for wide-field fluorescence and two-photon optical systems. **(A)** Schematic of recording system: An evenly-spaced 4×3 grid of sensors detects two sources. Sensed regions have pitch of $10 \mu\text{m}$, and neurons are separated on the x -axis by $20 \mu\text{m}$. **(B,E,H)** CRBs with a focal depth of $100 \mu\text{m}$. **(C,F,I)** CRBs with a focal depth of $200 \mu\text{m}$. **(D,G,J)** CRBs with a focal depth of $500 \mu\text{m}$. CRBs for the x , y , and z coordinates are in the first, second, and third rows, respectively, and are reported as standard deviations. See Table 2 for parameter values.

much is still experimentally unknown about noise sources and their relation to neural activity. While our demonstrations cannot give precise predictions about the capabilities of recording technologies, they demonstrate general scaling properties of the technologies, as well as illustrate situations in which the framework could be useful with more detailed models of neural recording.

A first simplification is that our demonstrations used approximate models of how neurons and noise affect sensor signals. Our demonstrations showed how we could use recording channels to identify the location of a fixed, known, activity. In practice these activities fluctuate over time, and can differ based on the type of neuron. In addition, we assumed that the effects of neural activity are linearly combined into the sensor signal. In practice, nonlinear effects such as sensor saturation may be important. Both can be incorporated into a Fisher information-based framework, although neither are treated here. Perhaps the largest simplifying model, the various noise sources were approximated by a simple function that ignores many potential sources of noise (see *Supplementary Information*). A comprehensive model of how noise affects neurons and sensors does not yet exist. Further research in this area will yield more informative results.

Second, we asked how we could use simplified models of recording systems to estimate the locations of neurons. For example, for optical recordings we assumed scattering through homogenous tissue, and for electrical recordings we ignored the filtering properties of electrodes. There exists a rich literature of modeling optical and electrical systems that could allow better models of recording modalities (e.g. [50, 60]); incorporating these models into the framework may alleviate some of the concerns over oversimplification, and may even provide a framework for validating those models.

In order to calculate the Fisher information contained by a given technique, we need to know its PSF and noise sources. When a technology is developed, experimentally determining these functions would allow this Fisher information to accurately be applied. These Fisher information calculations could determine how optimal a technique’s performance is. This information may then influence further design choices.

6 Additional Methods

6.1 Noise Calculations

In our Applications simulations, we make several assumptions about noise. We assume noise sources are uncorrelated (i.e. the noise from each neuron is independent and independently distributed). The sensor signal variance arises from signal dependent noise, with a standard deviation proportional to the mean signal. The signal dependent noise can be on all background neurons and/or on the sensor. As the mean activity is I_0 , the standard deviation of the activity is $\alpha \cdot I_0$, where α is a constant. The activity that reaches the sensor i (the signal) from a given neuron j then has a variance of $\sigma_j^2 = \alpha \cdot (I_0 \cdot w(\mathbf{d}^{i,j}))^2$. As the noise sources are independent, their variances can be added, so $\sigma_{noise}^2 = \sum_{j \neq k} \sigma_j^2$ (recall that we do not include noise from the neuron of interest). In simulations with two neurons of interest, we do not include noise from both neurons. We assume that neurons are uniformly distributed across the brain with density ρ_{space} and that all neurons have the same probability of firing at a given time, ρ_{fire} .

$$\begin{aligned} \sigma_{noise}^2 &= \alpha_{sens} \rho_{fire} \rho_{space} \int_V I_0^2 w^2 dV + \alpha_{neur} \rho_{fire} \rho_{space} \int_V I_0^2 w^2 dV \\ &= \alpha \rho_{fire} \rho_{space} \int_V I_0^2 w^2 dV \end{aligned} \quad (7)$$

In our simulations, we set $\alpha = 0.1$ (action potentials have SNRs ranging from 5-25 [61]), $\rho_{fire} = 0.01$ (assuming neurons on average fire at 5 Hz [62] and action potentials last ≈ 2 ms), and $\rho_{space} = 67000 \text{ mm}^{-3}$ (dividing the number of neurons in the human brain, $\approx 8 \times 10^{10}$ [63] by its volume, $\approx 1200 \text{ cm}^3$ [64]).

6.2 Applications: Electrode Grid Analysis

Electrode locations were assigned to nodes on a $1 \mu\text{m}$ grid spanning a $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ cube using the following procedures:

Columnar 6×6 : Column locations were spaced evenly, $200 \mu\text{m}$ apart, on a 6×6 grid in the x-y plane. 101 electrodes were distributed evenly along each column, $10 \mu\text{m}$ apart.

Columnar 10×10 : Column locations were spaced evenly, $111 \mu\text{m}$ apart, on a 10×10 grid in the x-y plane. 31 electrodes were distributed evenly along each column, $33 \mu\text{m}$ apart.

Random: Locations on the grid were drawn from a uniform random distribution with replacement.

Planar: Electrodes were placed on a uniform 61×61 grid in the x-y plane, corresponding to a grid spacing of $17 \mu\text{m}$, with a depth of $500 \mu\text{m}$.

Grid: Electrodes were placed on a uniform $15 \times 15 \times 15$ grid in the volume, corresponding to a grid spacing of $71 \mu\text{m}$.

These procedures give locations for 3636, 3100, 3636, 3721, and 3375 electrodes respectively.

For each arrangement of electrodes, we determine the 3×3 Fisher information matrix about the Cartesian location of a source contained by each electrode. We then sum these matrices under the assumption of independent noise on sensors, giving the aggregate Fisher information matrix for sensor i :

$$\mathcal{I}_{mn} = \frac{I_k^2}{\sigma_{noise}^2} \sum_{t=1}^e \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_m} \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_n} \quad (8)$$

Where I_k is the intensity of the neuron of interest, $\theta_m, \theta_n \in \{x, y, z\}$ of the given neuron are the parameters of interest, and e is the number of electrodes (See *Supplementary Information*). The Cramer-Rao bounds on x, y, z are the diagonal elements of \mathcal{I}^{-1} .

7 Acknowledgements

We would like to thank Dario Amodei and Darcy Peterka for their helpful comments. We would like to thank Dan Dombeck for helpful discussions regarding optics and Mikhail Shapiro for discussions regarding MR applications.

Thaddeus Cybulski, Joshua Glaser, and Bradley Zamft are supported by NIH grant 5R01MH103910. Adam Marblestone is supported by the Fannie and John Hertz Foundation fellowship. Konrad Kording is funded in part by the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust. Konrad Kording is also supported by NIH grants 5R01NS063399, P01NS044393, and 1R01NS074044. George Church

acknowledges support from the Office of Naval Research and the NIH Centers of Excellence in Genomic Science. Edward Boyden acknowledges funding by Allen Institute for Brain Science; AT&T; Google; IET A. F. Harvey Prize; MIT McGovern Institute and McGovern Institute Neurotechnology (MINT) Program; MIT Media Lab and Media Lab Consortia; New York Stem Cell Foundation-Robertson Investigator Award; NIH Director’s Pioneer Award 1DP1NS087724, NIH Transformative Awards 1R01MH103910 and 1R01GM104948, NSF INSPIRE Award CBET 1344219, Paul Allen Distinguished Investigator in Neuroscience Award; Skolkovo Institute of Science and Technology; Synthetic Intelligence Project (& its generous donors).

8 References

- [1] A. P. Alivisatos, A. M. Andrews, E. S. Boyden, M. Chun, G. M. Church, K. Deisseroth, J. P. Donoghue, S. E. Fraser, J. Lippincott-Schwartz, L. L. Looger, S. Masmanidis, P. L. McEuen, A. V. Nurmikko, H. Park, D. S. Peterka, C. Reid, M. L. Roukes, A. Scherer, M. Schnitzer, T. J. Sejnowski, K. L. Shepard, D. Tsao, G. Turrigiano, P. S. Weiss, C. Xu, R. Yuste, and X. Zhuang. Nanotools for neuroscience and brain activity mapping. *ACS Nano*, 7(3):1850–66, 2013.
- [2] Adam H Marblestone, Bradley M Zamft, Yael G Maguire, Mikhail G Shapiro, Thaddeus R Cybulski, Joshua I Glaser, Dario Amodei, Ben Stranges, Reza Kalhor, David A Dalrymple, Dongjin Seo, Elad Alon, Michel M Maharbiz, Jose M Carmena, Jan M Rabaey, Edward S Boyden, George M Church, and Konrad P Kording. Physical principles for scalable neural recording. *Front Comput Neurosci*, 7:137, 2013.
- [3] Eran a Mukamel, Axel Nimmerjahn, and Mark J Schnitzer. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron*, 63(6):747–60, September 2009.
- [4] AJ Den Dekker and A Van den Bos. Resolution: a survey. *JOSA A*, 14(3):547–557, 1997.
- [5] Yoichi Onodera, Yuji Kato, and Koichi Shimizu. Suppression of scattering effect using spatially dependent point spread function. In *Advances in Optical Imaging and Photon Migration*. Optical Society of America, 1998.
- [6] SB Colak, DG Papaioannou, GW t Hooft, MB Van der Mark, H Schomberg, JCJ Paasschens, JBM Melissen, and NAAJ Van As-

- ten. Tomographic image reconstruction from optical projections in light-diffusing media. *Applied optics*, 36(1):180–213, 1997.
- [7] Yan Yan and Gengsheng L Zeng. Scatter and blurring compensation in inhomogeneous media using a postprocessing method. *Journal of Biomedical Imaging*, 2008:15, 2008.
 - [8] Michael Broxton, Logan Groesenick, Samuel Yang, Noy Cohen, Aaron Andalman, Karl Deisseroth, and Marc Levoy. Wave optics theory and 3-d deconvolution for the light field microscope. *Optics express*, 21(21):25418–25439, 2013.
 - [9] Morteza Shahram and Peyman Milanfar. Imaging below the diffraction limit: A statistical analysis. *Image Processing, IEEE Transactions on*, 13(5):677–689, 2004.
 - [10] Morteza Shahram. *Statistical and information-theoretic analysis of resolution in imaging and array processing*. Thesis, 2005.
 - [11] Solomon Kullback. *Information theory and statistics*. Courier Dover Publications, 1997.
 - [12] H. Cramér. *Methods of mathematical statistics*. Princeton University Press, Princeton, 1946.
 - [13] Eftychios A Pnevmatikakis, Timothy A Machado, Logan Groesenick, Ben Poole, Joshua T Vogelstein, and Liam Paninski. Rank-penalized nonnegative spatiotemporal deconvolution and demixing of calcium imaging data.
 - [14] Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.
 - [15] Alison L Barth and James FA Poulet. Experimental evidence for sparse firing in the neocortex. *Trends Neurosci*, 35(6):345–355, 2012.
 - [16] Daniel J Denman and Diego Contreras. The structure of pairwise correlation in mouse primary visual cortex reveals functional organization in the absence of an orientation map. *Cerebral Cortex*, 2013.
 - [17] Jeremiah Y Cohen, Erin A Crowder, Richard P Heitz, Chenchal R Subraveti, Kirk G Thompson, Geoffrey F Woodman, and Jeffrey D Schall. Cooperation and competition among frontal eye field neurons during visual target selection. *The Journal of neuroscience*, 30(9):3227–3238, 2010.

- [18] Marlene R Cohen and Adam Kohn. Measuring and interpreting neuronal correlations. *Nat Neurosci*, 14(7):811–819, 2011.
- [19] Wyeth Bair, Ehud Zohary, and William T Newsome. Correlated firing in macaque visual area mt: time scales and relationship to behavior. *The Journal of neuroscience*, 21(5):1676–1697, 2001.
- [20] Daniel Johnston, Samuel Miao-Sin Wu, and Richard Gray. *Foundations of cellular neurophysiology*. MIT press Cambridge, 1995.
- [21] Christoph Cremer and Barry R Masters. Resolution enhancement techniques in microscopy. *The European Physical Journal H*, pages 1–64, 2013.
- [22] Sean Quirin, Sri Rama Prasanna Pavani, and Rafael Piestun. Optimal 3d single-molecule localization for superresolution microscopy with aberrations and engineered point spread functions. *Proceedings of the National Academy of Sciences*, 109(3):675–679, 2012.
- [23] Raimund J Ober, Sripad Ram, and E Sally Ward. Localization accuracy in single-molecule microscopy. *Biophys J*, 86(2):1185–1200, 2004.
- [24] Franois Aguet, Dimitri Van De Ville, and Michael Unser. A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles. *Opt. Express*, 13(26):10503–10522, 2005.
- [25] Kim A Winick. Cramer-rao lower bounds on the performance of charge-coupled-device optical position estimators. *JOSA A*, 3(11):1809–1815, 1986.
- [26] Morteza Shahram and Peyman Milanfar. Statistical and information-theoretic analysis of resolution in imaging. *Information Theory, IEEE Transactions on*, 52(8):3411–3437, 2006.
- [27] Joo Sanches, Ins Sousa, and Patricia Figueiredo. Bayesian fisher information criterion for sampling optimization in asl-mri. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 880–883. IEEE.
- [28] Edwin A Marengo, Maytee Zambrano-Nunez, and David Brady. Cramer-rao study of one-dimensional scattering systems: Part i: Formulation. In *Proceedings of the 6th IASTED International Conference on Antennas, Radar, and Wave Propagation (ARP'09)*, pages 1–8.
- [29] Carl W Helstrom. Detection and resolution of incoherent objects by a background-limited optical system. *JOSA*, 59(2):164–175, 1969.

- [30] Eran A. Mukamel and Mark J. Schnitzer. Unified Resolution Bounds for Conventional and Stochastic Localization Fluorescence Microscopy. *Physical Review Letters*, 109(16):168102, October 2012.
- [31] Yoav Shechtman, Steffen J. Sahl, Adam S. Backer, and W.E. Moerner. Optimal Point Spread Function Design for 3D Imaging. *Physical Review Letters*, 113(13):133902, September 2014.
- [32] Stefanie Hampel, Phuong Chung, Claire E McKellar, Donald Hall, Loren L Looger, and Julie H Simpson. Drosophila brainbow: a recombinase-based fluorescence labeling technique to subdivide neural expression patterns. *Nat Methods*, 8(3):253–259, 2011.
- [33] K. P. Kording. Of toasters and molecular ticker tapes. *PLoS Comput Biol*, 7(12):e1002291, 2011.
- [34] Anthony M Zador, Joshua Dubnau, Hassana K Oyibo, Huiqing Zhan, Gang Cao, and Ian D Peikon. Sequencing the connectome. *PLoS Biol*, 10(10):e1001411, 2012.
- [35] Harry L Van Trees. *Detection, estimation, and modulation theory*. John Wiley & Sons, 2004.
- [36] Justin Dauwels. Computing bayesian cramer-rao bounds. In *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pages 425–429. IEEE, 2005.
- [37] John D Gorman and Alfred O Hero. Lower bounds for parametric estimation with constraints. *Information Theory, IEEE Transactions on*, 36(6):1285–1301, 1990.
- [38] Charles Matson and Alim Haji. Biased cramer-rao lower bound calculations for inequality-constrained estimators (preprint). Technical report, DTIC Document, 2006.
- [39] Misha B. Ahrens, Michael B. Orger, Drew N. Robson, Jennifer M. Li, and Philipp J. Keller. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nat Meth*, 10(5):413–420, 2013.
- [40] Robert Prevedel, Y-G Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrdel, Ramesh Raskar, Manuel Zimmer, and Edward S Boyden. Simultaneous whole-animal 3d-imaging of neuronal activity using light field microscopy. *arXiv preprint arXiv:1401.5333*, 2014.
- [41] Dongjin Seo, Jose M Carmena, Jan M Rabaey, Elad Alon, and Michel M Maharbiz. Neural dust: An ultrasonic, low power solution

- for chronic brain-machine interfaces. *arXiv preprint arXiv:1307.2196*, 2013.
- [42] Ho-Chul Shin, Richard Prager, James Ng, Henry Gomersall, Nick Kingsbury, Graham Treece, and Andrew Gee. Sensitivity to point-spread function parameters in medical ultrasound image deconvolution. *Ultrasonics*, 49(3):344–357, 2009.
 - [43] Jrgen Arendt Jensen. A model for the propagation and scattering of ultrasound in tissue. *The Journal of the Acoustical Society of America*, 89:182, 1991.
 - [44] Qin Qin. Point spread functions of the t2 decay in k-space trajectories with long echo train. *Magnetic Resonance Imaging*, 2012.
 - [45] Robert Carl Smith and Robert C Lange. *Understanding magnetic resonance imaging*. CRC Press, 1998.
 - [46] Christoph J Engelbrecht and Ernst H Stelzer. Resolution enhancement in a light-sheet-based microscope (spim). *Optics letters*, 31(10):1477–1479, 2006.
 - [47] R. Segev, J. Goodhouse, J. Puchalla, and 2nd Berry, M. J. Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nat Neurosci*, 7(10):1154–61, 2004.
 - [48] C. M. Gray, P. E. Maldonado, M. Wilson, and B. McNaughton. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods*, 63(1-2):43–54, 1995.
 - [49] Johann Heinrich Lambert and E. Anding. *Lamberts Photometrie. (Photometria, sive De mensura et gradibus luminis, colorum et umbrae) (1760)*. Ostwalds Klassiker der exakten Wissenschaften. W. Engelmann, Leipzig,, 1892.
 - [50] P. Theer and W. Denk. On the fundamental imaging-depth limit in two-photon microscopy. *J Opt Soc Am A Opt Image Sci Vis*, 23(12):3139–49, 2006.
 - [51] P. Tian, A. Devor, S. Sakadzic, A. M. Dale, and D. A. Boas. Monte carlo simulation of the spatial resolution and depth sensitivity of two-dimensional optical imaging of the brain. *J Biomed Opt*, 16(1):016006, 2011.

- [52] George Biddell Airy. On the diffraction of an object-glass with circular aperture. *Transactions of the Cambridge Philosophical Society*, 5:283, 1835.
- [53] D Thomann, DR Rines, PK Sorger, and G Danuser. Automatic fluorescent tag detection in 3d with superresolution: application to the analysis of chromosome movement. *Journal of microscopy*, 208(1):49–64, 2002.
- [54] Hagai Kirshner, Franois Aguet, Daniel Sage, and Michael Unser. 3d psf fitting for fluorescence microscopy: implementation and localization application. *Journal of microscopy*, 249(1):13–25, 2013.
- [55] Jos RA Torreao and Joao L Fernandes. Single-image shape from defocus. In *Computer Graphics and Image Processing, 2005. SIBGRAPI 2005. 18th Brazilian Symposium on*, pages 241–246. IEEE.
- [56] F. Helmchen and W. Denk. Deep tissue two-photon microscopy. *Nat Methods*, 2(12):932–40, 2005.
- [57] Harry S Orbach and Lawrence B Cohen. Optical monitoring of activity from many areas of the in vitro and in vivo salamander olfactory bulb: a new method for studying functional organization in the vertebrate central nervous system. *The Journal of neuroscience*, 3(11):2251–2262, 1983.
- [58] Emmanuelle Chaigneau, Amanda J Wright, Simon P Poland, John M Girkin, and R Angus Silver. Impact of wavefront distortion and scattering on 2-photon microscopy in mammalian brain tissue. *Optics express*, 19(23):22755, 2011.
- [59] Anthony N Zorzos, Jorg Scholvin, Edward S Boyden, and Clifton G Fonstad. Three-dimensional multiwaveguide probe array for light delivery to distributed brain circuits. *Optics letters*, 37(23):4841–4843, 2012.
- [60] Luis A Camuas-Mesa and Rodrigo Quian Quiroga. A detailed and fast model of extracellular recordings. *Neural Comput*, 25(5):1191–1212, 2013.
- [61] Jonathan Erickson, Angela Tooker, Y-C Tai, and Jerome Pine. Caged neuron mea: A system for long-term investigation of cultured neural network connectivity. *Journal of neuroscience methods*, 175(1):1–16, 2008.
- [62] Julia J Harris, Renaud Jolivet, and David Attwell. Synaptic energy use and supply. *Neuron*, 75(5):762–777, 2012.

- [63] Frederico AC Azevedo, Ludmila RB Carvalho, Lea T Grinberg, Jos Marcelo Farfel, Renata EL Ferretti, Renata EP Leite, Roberto Lent, and Suzana HerculanoHouzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009.
- [64] John S Allen, Hanna Damasio, and Thomas J Grabowski. Normal neuroanatomical variation in the human brain: An mrivolumetric study. *American journal of physical anthropology*, 118(4):341–358, 2002.
- [65] Robert Tyson. *Adaptive optics engineering handbook*, volume 67. CRC Press, 1999.
- [66] Alan C Bovik. *Handbook of image and video processing*. Access Online via Elsevier, 2010.

9 Supplementary Information

9.1 Noise Sources

The Fisher information framework allows for arbitrary noise sources, so long as they are able to be modeled. However, to demonstrate potential applications, we used a very simplified noise model that only considered signal dependent noise where the standard deviation was proportional to the mean.

There are multiple potentially relevant sources of noise that could readily be included in our model. (1) Each sensor has a constant level of noise simply due to thermal effects. (2) Many sensors have an additional variance that is proportional to the square of the signals, e.g. reference fluctuations. (3) Many sensors have an additional variance that is proportional to the signal, e.g. due to low numbers of photons (shot noise). (4,5) Each neuron may produce constant noise, e.g. background fluorescence of dyes. These neural noise sources may be independent or correlated. (6,7) Each neuron may produce variance that quadratically depends on its activation, e.g. action potentials that propagate back into varying parts of the dendritic tree.⁶ These neural noise sources may be independent or correlated. (8,9) Each

⁶In a simplistic model, when a neuron fires, the action potential spreads into some variable proportion of the dendritic tree. If the recorded signal is dependent on the proportion of dendritic branches the action potential propagates into, then the standard deviation of the recorded signal is proportionate to the mean signal entering the dendrites.

neuron may produce variance that linearly depends on its signal strength, e.g. fluorophore activations. These neural noise sources may be independent or correlated. We have some knowledge about the exact sizes of these signals [2], but most of these numbers are hard to know. They may be reasonable to measure in future experiments.

Taking these signals together, we obtain the following noise level on a sensor i (given a recording of N firing neurons indexed by j):

$$\begin{aligned}
\sigma_{noise}^2 = & \underbrace{\sigma_{sens}^2}_{\text{constant sensor noise}} + \underbrace{\alpha_1 \sum_{j=1}^N I_0^2(w(\mathbf{d}^{i,j}))^2}_{\text{ind sensor } \sigma \propto \mu \text{ SDN}} + \underbrace{\alpha_2 \sum_{j=1}^N I_0(w(\mathbf{d}^{i,j}))^2}_{\text{ind sensor } \sigma^2 \propto \mu \text{ SDN}} \\
& + \underbrace{\alpha_3 N \cdot \sigma_{neur}^2}_{\text{constant ind neuron noise}} + \underbrace{\alpha_4 N \cdot \sigma_{neur}^2}_{\text{constant corr neuron noise}} \\
& + \underbrace{\alpha_5 \sum_{j=1}^N I_0^2(w(\mathbf{d}^{i,j}))^2}_{\text{ind neuron } \sigma \propto \mu \text{ SDN}} + \underbrace{\alpha_6 \left(\sum_{j=1}^N I_0(w(\mathbf{d}^{i,j})) \right)^2}_{\text{corr neuron } \sigma \propto \mu \text{ SDN}} \\
& + \underbrace{\alpha_7 \sum_{j=1}^N I_0(w(\mathbf{d}^{i,j}))}_{\text{ind neuron } \sigma^2 \propto \mu \text{ SDN}} + \underbrace{\alpha_8 \left(\sum_{j=1}^N \sqrt{I_0(w(\mathbf{d}^{i,j}))} \right)^2}_{\text{corr neuron } \sigma^2 \propto \mu \text{ SDN}}
\end{aligned} \tag{9}$$

where *ind* and *corr* refer to independent and correlated noise sources, and *SDN* refers to signal dependent noise.

Assuming, as we do in the main text's demonstrations, that neurons are uniformly distributed and have a uniform firing rate across the entire volume:

$$\begin{aligned}
\sigma_{noise}^2 = & \underbrace{\sigma_{sens}^2}_{\text{constant sensor noise}} + \underbrace{\alpha_1 \rho_{fire} \rho_{space} \int I_0^2 w^2 dV}_{\text{ind sensor } \sigma \propto \mu \text{ SDN}} + \underbrace{\alpha_2 \rho_{fire} \rho_{space} \int I_0 w dV}_{\text{ind sensor } \sigma^2 \propto \mu \text{ SDN}} \\
& + \underbrace{\alpha_3 \rho_{fire} \rho_{space} V \cdot \sigma_{neur}^2}_{\text{constant ind neuron noise}} + \underbrace{\alpha_4 \rho_{fire} \rho_{space} V \cdot \sigma_{neur}^2}_{\text{constant corr neuron noise}} \\
& + \underbrace{\alpha_5 \rho_{fire} \rho_{space} \int I_0^2 w^2 dV}_{\text{ind neuron } \sigma \propto \mu \text{ SDN}} + \underbrace{\alpha_6 \rho_{fire} \rho_{space} \left(\int I_0 w dV \right)^2}_{\text{corr neuron } \sigma \propto \mu \text{ SDN}} \\
& + \underbrace{\alpha_7 \rho_{fire} \rho_{space} \int I_0 w dV}_{\text{ind neuron } \sigma^2 \propto \mu \text{ SDN}} + \underbrace{\alpha_8 \rho_{fire} \rho_{space} \left(\int \sqrt{I_0 w} dV \right)^2}_{\text{corr neuron } \sigma^2 \propto \mu \text{ SDN}}
\end{aligned} \tag{10}$$

Both constant and shot noise terms can be minimized in their effect by optimizing the experimental design, e.g. through good dyes and strong illumination (but see [2]).

In addition, in the main text we assume that the noise is Gaussian, which has also been assumed in previous statistical formulations [10, 26]. This assumption has been shown to be valid for thermal noise and shot noise in some conditions [65, 66].

9.2 Fisher Information Derivation

We have a distribution $f(X_i; \boldsymbol{\theta}) = \mathcal{N}\left(\sum_j I_j w(\mathbf{d}^{i,j}), \sigma_{noise}^2\right)$ (See (5b)).

Focusing on a particular neuron with index k , this equation becomes

$$f_i(X; \mathbf{d}) = \mathcal{N}\left(\left(I_k w(\mathbf{d}^{i,k}) + \sum_{j \neq k} I_j w(\mathbf{d}^{i,j})\right), \sigma_{noise}^2\right). \tag{11}$$

We are interested in finding its Fisher information about the parameter vector $\boldsymbol{\theta}$ that contains the directions of interest. Let θ_m be $d_p^{i,k}$ and θ_n be $d_q^{i,k}$, where $d_p^{i,k}$ and $d_q^{i,k}$ are two different location parameters of neuron k with respect to sensor i . For simplicity of notation, we let $B = \sum_{j \neq k} I_j w(\mathbf{d}^{i,j})$.

$$\begin{aligned}
\mathcal{I}_{mn} &= E \left[\left(\frac{\partial}{\partial \theta_m} \ln(f(X; \boldsymbol{\theta})) \right) \left(\frac{\partial}{\partial \theta_n} \ln(f(X; \boldsymbol{\theta})) \right) \right] \\
&= E \left[\left(\frac{1}{f(X; \boldsymbol{\theta})} \frac{\partial}{\partial \theta_m} f(X; \boldsymbol{\theta}) \right) \left(\frac{1}{f(X; \boldsymbol{\theta})} \frac{\partial}{\partial \theta_n} f(X; \boldsymbol{\theta}) \right) \right]
\end{aligned} \tag{12}$$

$$\begin{aligned}
\frac{\partial}{\partial \theta_m} \ln(f(X; \boldsymbol{\theta})) &= \frac{1}{f(X; \boldsymbol{\theta})} \frac{\partial}{\partial \theta_m} f(X; \boldsymbol{\theta}) \\
&= \frac{1}{\frac{1}{\sqrt{2\pi\sigma_{noise}^2}} \exp\left(\frac{-(X - (I_k w(\mathbf{d}^{i,k}) + B))^2}{2\sigma_{noise}^2}\right)} \\
&\quad \frac{1}{\sqrt{2\pi\sigma_{noise}^2}} \exp\left(\frac{-(X - (I_k w(\mathbf{d}^{i,k}) + B))^2}{2\sigma_{noise}^2}\right) \\
&\quad \frac{\partial}{\partial \theta_m} \left(\frac{-(X - (I_k w(\mathbf{d}^{i,k}) + B))^2}{2\sigma_{noise}^2} \right) \\
&= \frac{I_k}{2\sigma_{noise}^2} 2 \left(X - (I_k w(\mathbf{d}^{i,k}) + B) \right) \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_m} \\
&= \frac{I_k}{\sigma_{noise}^2} \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_m} \left(X - (I_k w(\mathbf{d}^{i,k}) + B) \right)
\end{aligned} \tag{13}$$

$$\begin{aligned}
\mathcal{I}_{mn} &= E \left[\left(\frac{I_k^2}{\sigma_{noise}^4} \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_m} \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_n} \left(X - (I_k w(\mathbf{d}^{i,k}) + B) \right)^2 \right) \right] \\
&= \frac{I_k^2}{\sigma_{noise}^4} \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_m} \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_n} E \left[\left(X - (I_k w(\mathbf{d}^{i,k}) + B) \right)^2 \right] \\
&= \frac{I_k^2}{\sigma_{noise}^2} \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_m} \frac{\partial w(\mathbf{d}^{i,k})}{\partial \theta_n}
\end{aligned} \tag{14}$$

The above derivation will hold for any distribution with zero-mean Gaussian noise.